AUTOMATIC MODELLING IMAGE REPRESENTED OBJECTS USING A STATISTIC BASED APPROACH

Maria João M. Vasconcelos and João Manuel R. S. Tavares
Faculdade de Engenharia da Universidade do Porto, Departamento de Engenharia Mecânica e Gestão Industrial
Instituto de Engenharia Mecânica e Gestão Industrial, Laboratório de Óptica e Mecânica Experimental
Rua Drº Roberto Frias s/n, 4200-465 Porto - PORTUGAL
Emails: merido@netcabo.pt; tavares@fe.up.pt

KEYWORDS
Computational Vision, Image Analysis, Segmentation, Recognition, Point Distribution Models, Active Shape Models, Active Appearance Models, Automatic Extraction of Landmarks.

ABSTRACT

This paper presents new methodologies to automatically extract significant points, from an object represented in images, useful to construct Point Distribution Models. Each model consists of a flexible shape template, describing how significant points of the object can vary, and a statistical model of the expected grey levels in regions around each model point. This information can be used to search objects in new images: Active Shape and Active Appearance Models. Both use PDMs for image analysis, to locate structures modeled in new images, or in a classifier, an estimate can be made of how likely the example in cause is a member of the class of shapes described by the model build. We present results for two objects: a hand and a face.

INTRODUCTION

One of the recent areas of interest in Computacional Vision is image analysis based on flexible models. These deformable models were developed to deal with problems where the images contained objects of variable shapes.

The methodology used in this paper, Point Distribution Models (PDMs), was initially proposed by (Cootes et al., 1992). The model is build based on a training set of images containing the object with different forms, and then the shape in each image is represented by a set of labelled points, known as landmarks. The point distribution model is obtained by capturing the statistics of the co-ordinates of the landmarks: after the sets are aligned, a Principal Component Analysis is made. A PDM is obtained with a small number of parameters linearly independent which translate the mean shape of the object in study as well as the main modes of variation.

In (Cootes and Taylor, 1992a), the authors describe the Active Shape Models (ASMs), an iterative technique for fitting flexible models to images objects. The technique is an iterative optimisation scheme for PDMs allowing initial estimates of the pose, scale and shape of an object in an image to be refined. The incorporation of statistical models of grey-level appearance in Active Shape Model search lead to improved reliability and accuracy (Cootes and Taylor, 1993). Grey-level appearance is represented by statistical models of the grey-levels in regions around each of the shape model points.

Another method for image analysis that uses the principles of the PDMs is the Active Appearance Model (AAM) (Cootes et al., 1998). The appearance model contains a statistical model of the shape and grey-level appearance of the object of interest. The search method starts to learn the relationship between model parameter displacements and the residual errors, induced between a training image and a synthesised model example during a training phase. To match the model in an image, AAM measure the current residuals and use the model to predict changes to the current parameters, leading to a better fit.

These statistical models have been very usefull for image analysis in different applications of computational vision. For instance, they can be used on several areas like: medicine, for locating bones and organs in medical images; industry, for industrial inspection; and security, for face recognition.

The annotation of the landmarks to build PDMs is the most time consuming step of the construction of these models, because is generally manually made. Some authors, like (Hill and Taylor, 1994; Baker and Matthews, 2002; Hicks et al., 2002; Angelopoulou and Psarrou, 2004), have been developing methods to automatize this stage. In this paper, we present new methods to automatically extract the landmarks to be considered in the modelling of objects, like hands and faces.

In next section, we explain how to build PDMs, ASMs and AAMs; and then, we present our methods to automatically extract the landmarks of objects like hands and faces; after this, some applications of these models are presented; and finally, in the last section, some conclusions and suggestions of future work are made.

POINT DISTRIBUTION MODEL

(Cootes et al., 1992) describe how to build flexible shape models called Point Distribution Models. These are generated from examples of shapes, where each shape is represented by a set of labelled landmark points. Usually, the landmarks represent the boundary or significant internal locations of an object (Figure 1).

After extracting the landmaks, all the training examples are aligned into a standard co-ordinate frame and a Principal Component Analysis is applied to the co-ordinates of the points. This produces the mean position for each landmark,

and a description of the main ways in which the points tend to move together.

Figure 1: Training image, landmarks and an image labelled with the landmark points (from left to right).

The equation below represents the Point Distribution Model or Shape Model and can be used to generate new shapes of the object:

$$x = \bar{x} + P_s b_s , \qquad (1)$$

where x represents the n points of the shape:

$$x = (x_0, x_1, \ldots, x_{n-1}, y_0, y_1, \ldots, y_{n-1}) ,$$

and (x_k, y_k) is the position of point k, $\bar{x}$ is the mean position of the points, $P_s = (p_{s1} \quad p_{s2} \quad \ldots \quad p_{st})$ is the matrix of the first t modes of variation, p_{si}, corresponding to the most significant eigenvectors in a Principal Component Analysis of the position variables, and $b_s = (b_{s1} \quad b_{s2} \quad \ldots \quad b_{st})^T$ is a vector of weigths for each mode.

If the shape parameters b are chosen inside suitable limits (derived from the training set), then the shapes generated by equation (1) will be similar to those given in the original training set.

The local grey-level environment about each landmark can also be modelled. Thus, statistical information is obtained about the mean and covariances of the grey values of the pixels around each landmark. This information can be used later in Active Shape Models to evaluate the match between landmarks or in Active Appearance Models to construct the appearance models, as we explain next.

Active Shape Model

After obtain the PDM and the grey level profiles for each landmark, we can search for the object modelled in new images using the Active Shape Models, an iterative technique for fitting flexible models to images objects (Cootes and Taylor, 1992a).

The technique is an iterative optimisation squeme for PDMs allowing initial estimates of the pose, scale and shape of an object in an image to be refined. The used approach is summarized on the following steps: 1) at each point in the models is calculated a suggested movement required to displace the point to a better position; 2) the changes in the overall position, orientation and scale of the model which best satisfy the displacements are calculated; 3) finally, any residual differences are used to deform the shape of the model object by calculating the required adjustments to the shape parameters.

In (Cootes et al., 1994) is presented an improvement for the active shape models, which uses multiresolution. So, initially the method constructs a multirresolution pyramid of the images, by applying a Gaussian mask, and study grey level profiles for the various levels. This makes active models faster and reliable.

Active Appearance Model

This approach was presented by (Cootes et al., 1998) and allow to construct texture and appearance models. These models are generated by combining a model of shape variation, with a model of the appearance variations in a shape-normalised frame. The statistical model of the shape was already described above in equation (1). To build a statistical model of the grey level appearance, we deform each example image so that its control points match the mean shape, by using a triangulation algorithm. We then sample the grey level information, g_{im} from the shape-normalised image over the region covered by the mean shape. To minimize the effect of global lighting variation, we normalize this vector, obtaining g. By applying a Principal Component Analysis to this data, we obtain a linear model, the texture model:

$$g = \bar{g} + P_g b_g , \qquad (2)$$

where $\bar{g}$ is the mean normalised grey level vector, P_g is a set of orthogonal modes of grey level variation and b_g is a set of grey level model parameters.

Therefore, the shape and appearance of any example can be summarized by the vectors b_s and b_g. Since there may be correlations between the shape and grey levels variations, we apply a further Principal Component Analysis to the data as follows. For each example we generate the concatenated vector:

$$b = \begin{pmatrix} W_s b_s \\ b_g \end{pmatrix} = \begin{pmatrix} W_s P_s^T (x - \bar{x}) \\ P_g^T (g - \bar{g}) \end{pmatrix}, \qquad (3)$$

where W_s is a diagonal matrix of weights for each shape parameter, allowing for the difference in units between the shape and the grey models. We apply a Principal Component Analysis on these vectors, giving a further model:

$$b = Qc , \qquad (4)$$

where Q are the eigenvectors of b, and c is the vector of appearance parameters controlling both the shape and the grey levels of the model. In this away, an example image can be synthecised for a given c by generating the shape-free grey level image, from the vector g, and deforming it using the control points described by x.

AUTOMATIC EXTRACTION OF LANDMARKS

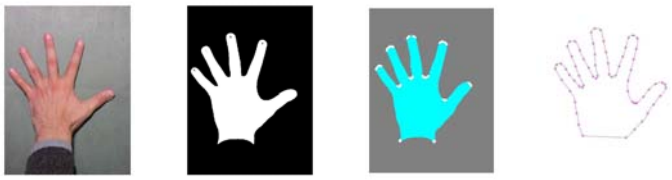
In this section we present four new methods which can be used to automatically extract landmaks from objects like hands and faces. In this work, the first one is used on hands and the next three on faces.

Hand labeling

The algorithm that automatically extracts landmarks from hands was constructed based on the methods proposed in (Angelopoulou and Psarrou, 2004; Carvalho and Tavares, 2005).

Our approach can be summarized on the following steps: 1) segmentation of the training image in order to isolate the object in study, by using an algorithm to detect skin regions (Carvalho and Tavares, 2005); 2) extract hand contour and find contour zones with high curvature, by using the k-curvature method (Lim, 1990); 3) find the contour corresponding to the hand, by deleting possible pulse and arm contours; 4) finally, choose the number of landmarks to considerer on high curvature zones and also the number of landmarks between them.

In figure 2, a) we present an example of a training image, in b) the segmentation resulted, in c) the contour found with the high curvature zones identified, and in d) the landmarks extracted (an example).



a) b) c) d)

Figure 2: a) Training image, b) segmentation result using the skin algorithm, c) hand contour and high curvature zones (white) found, and d) landmarks extracted.

Face contour extraction

This method extract significant points from face images, like: chin, eye, eyebrow or mouth landmarks.

The first step of our method uses, as the method for hand extraction, the skin detection algorithm to localize the face region. In Figure 3, is presented an example of a segmentation result in localizing skin regions in an image.

Studies like (Campadelli et al., 2003) show that the use of chrominance maps are usefull for eyebrows and eyes localization. Chromatic colors can be obtained from RGB space by the transformation:

$$Cr = \frac{R}{R+G+B}, \quad Cb = \frac{B}{R+G+B} \qquad (5)$$

Eyes are characterized in $CbCr$ plane by low values on the red component, Cr, and high values on the blue component, Cb, so the chrominance map for eyes can be defined by the following equation:

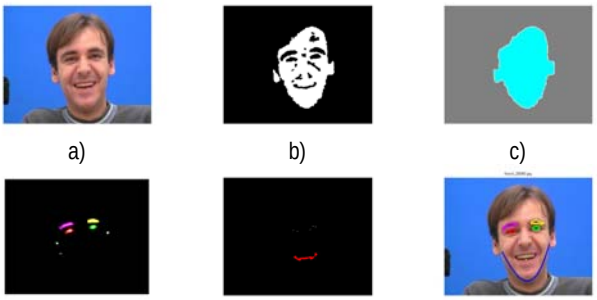
$$EyeMap = \frac{1}{3}\left\{\left(Cb^2\right) + \left(\hat{Cr}\right)^2 + \left(\frac{Cb}{Cr}\right)\right\}, \qquad (6)$$

where Cb^2, $\hat{Cr}^2$ and Cb/Cr are normalized to the range $[0,255]$ and $\hat{Cr}$ is the negative of Cr (ie, $\hat{Cr} = 255 - Cr$). In our work, the EyeMap is used also to identify the eyebrows region.

In other hand, the mouth region is identified using the HSV space, where H, S, V represent hue, saturation and value, respectively, where mouth is caracterized by having high values on the saturation component.

By congregating the localization of face, eyebrows, eyes and mouth contours is possible to extract landmarks from each of these zones. Considering that the zone of the chin is the most important in the face contour, we only use the inferior contour between ears.

In Figure 3 some results of this method are present for a training image example.



a) b) c) d) e) f)

Figure 3: a) Training image, b) segmentation result using the skin algorithm, c) face contour extracted, d) eyebrows and eyes found, e) mouth identified, and f) final contours obtained.

Face regular mesh

The second method developed for automatically landmark extraction from face images is based on the worked presented in (Baker and Matthews, 2004) that, to construct active appearance models, consider landmarks as the nodes of a mesh.

Our method starts to identify face and eye regions like described in the last section, and adjust a regular rectangular mesh to the face region, rotating it according the angle given by eye centroids. The nodes of the mesh are considered as landmarks and used for active appearance models.

Figure 4 shows the face mesh result of this method in a training image example.

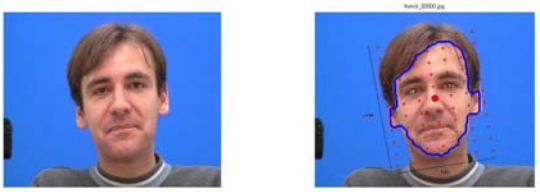
Face adaptative multiresolution mesh

Finally, our third method combines the philosophy of the method presented in section 3.2, using face, eyes and mouth localization, and of the method of section 3.3, considering the landmarks as the nodes of the used meshes. So, our new method builds a multiresolution mesh considering face, eyes and mouth positions.

After localizing face, eyes and mouth regions like indicated in section 3.2, this method constructs two adaptative meshes, in the eye and mouth regions acording their localization, and

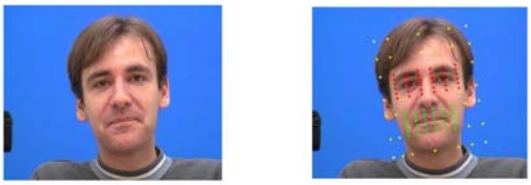
adds additional nodes, in the large mesh (that contains the face region), defined by the external edges and the bounds of the sub-meshes used in the regions of eyes and the mouth.

One example of the resulting final mesh using our method is presented in Figure 5.



a) b)

Figure 4: a) Training image, b) face regular mesh (red points) adapted to the face region (face contour in blue) and rotated according to the eyes direction (yellow).



a) b)

Figure 5: a) Training image and b) example of an adaptative multiresolution mesh obtained for a face.

In all the methods presented, we can choose the parameters that define the resulting mesh; that is, the number of lines and columns.

RESULTS

The methodologies described have been used in this work to generate shape models and appearance models for two classes of objects: hands and faces.

To generate the shape models, it was developed by us some implementations in MATLAB, based on the software *Active Shape Models* (Hamarneh, 1999). In other side, to study the appearance models, we used the *Modelling and Search Software* available in (Cootes, 2004). The sets of images considered in this study are available in (Stegmann and Gomez, 2002; Cootes, 2004).

For the hand models, we used a training set of 25 images, each one automatically labelled with 65 landmarks around its boundary (Figure 2d). Then a shape model was trained on this data, and it was found that 95% of the shape variance could be explained just by the first 6 modes. The first three modes are shown in Figure 6, and consist in the combination of global transformations and movements of the fingers.

Figure 7 presents the segmentation results for the active shape model in a test image. The grey levels profile used for the construction of active shape model were considered with 7 pixels long. The search runs at most 10 iterations on each resolution level, starting with a level 3 image.

In the case of the face models, a training set of 22 images was used, the face contour model extracted 44 landmarks around the main features, the regular face mesh extract 36 landmarks, and the adaptative multiresolution face mesh model used 54 landmarks.

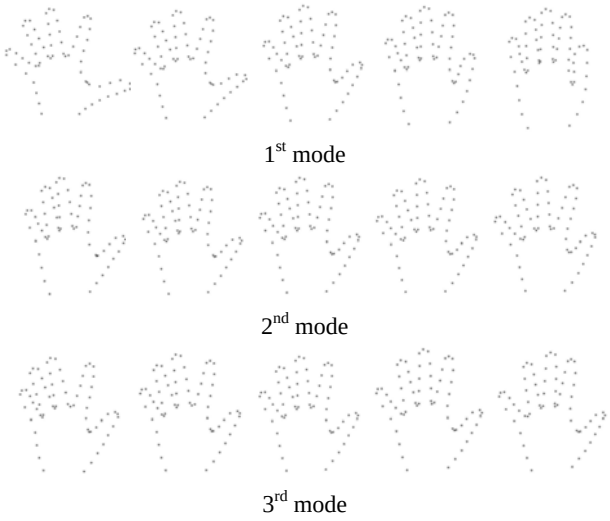


Figure 6: Effects of varying each of the first three parameters of the hand shape model individually ($\pm 2sd$).

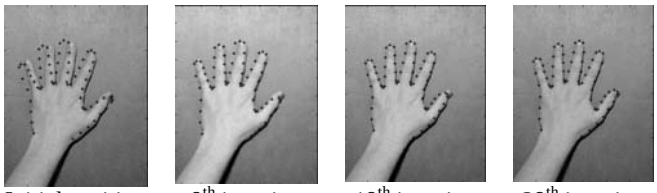


Figure 7: Test image with initial position of the mean model overlapped, and after the 9th, 19th and 29th iteration of the search with the active shape model built.

For the first face shape model trained (face contour), it was found that for 95% of the shape variance could be explained only by the first 13 modes of variation. By other hand, for the texture model it was found that 95% of the variance could be explained by the first 15 modes of variation. Finally, the appearance model needs only 12 modes of variation to explain 95% of the observed variance; the first four modes of appearance variation are shown in Figure 8.

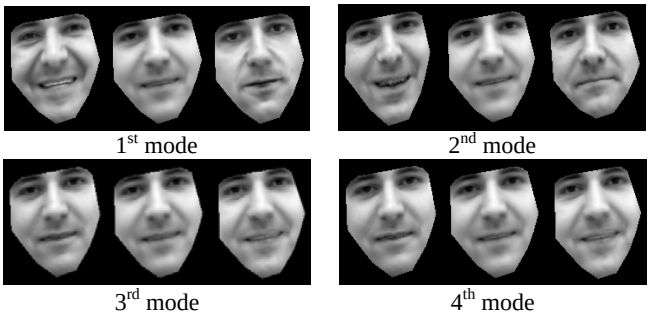


Figure 8: First four modes of appearance variation for the contour face model built ($\pm 2sd$).

For the second face shape model trained (using a regular face mesh), it was found that for 95% of the variance could be explained only by the first mode of variation, Figure 9. In other hand, for the texture model, it was found that 95% of the variance could be explained by the first 11 modes of variation. Finally, the appearance model needs only 5 modes

of variation to explain 95% of the observed variance; the first two modes of variation of the appearance model are shown in Figure 9.

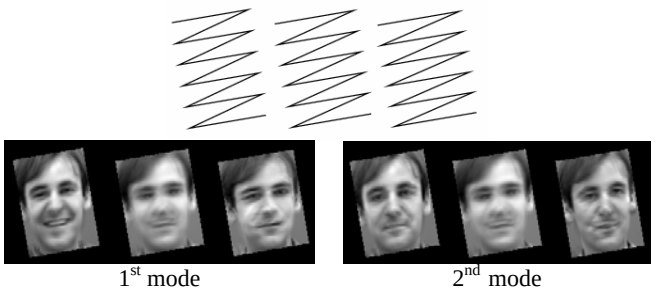


Figure 9: First mode of shape variation and the two first modes of appearance variation for the regular face mesh model built ($\pm 2sd$). (In the representation above, the nodes of the used mesh are linked by "extra" lines just for visualization purpose.)

For the third and last face shape model trained (using an adaptative multiresolution face mesh), it was found that for 95% of the variance could be explained only by the first 3 modes of variation. In the other hand, for the texture model, it was found that 95% of the variance could be explained by the first 14 modes of variation. In last, the appearance model needs only 8 modes of variation to explain 95% of the observed variance; the first four modes of variation of the appearance model are shown in Figure 10.

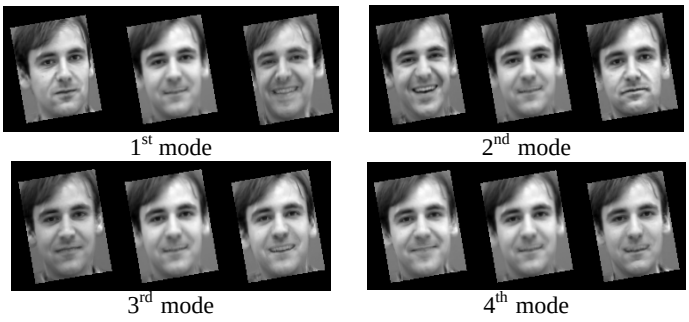


Figure 10: First four modes of appearance variation for the adaptative multiresolution face mesh model considered ($\pm 2sd$).

In Fig. 11, 12 and 13 are presented some segmentation results for the active appearance models, considered above, in a test image. In active appearance search, 5 levels of resolution were used and a maximum of 5 iterations were allowed per level.

CONCLUSIONS

The methods described in this paper, allow the automatic building of flexible models, for deformable objects represented in example sets of images, easily.

Both active shape models and active appearance models built in this work, obtained good results in the recognition of the objects modelled in new images. Is important to refer that the test images considered in the previous section did not belong to the respective training set.

The methods presented to automatically extract the landmarks of objects, of type hand and face, shown to be reliable and allow the building of active shape models and active appearance models in a full automatic way.

As future work, we are intended to experiment and compare the models automatically build using our methodology in different kinds of applications. One of the possibilities is the segmentation of objects represented in medical images.

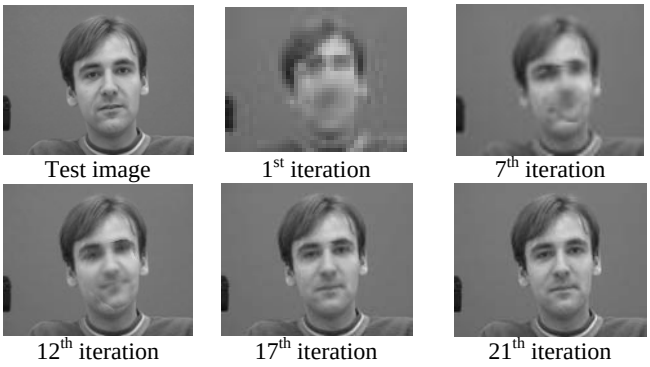


Figure 11: Test image with initial position of the mean model overlapped, and after the 1st, 7th, 12th, 17th and 21th iteration of the search with the active appearance model built for the face contour model.

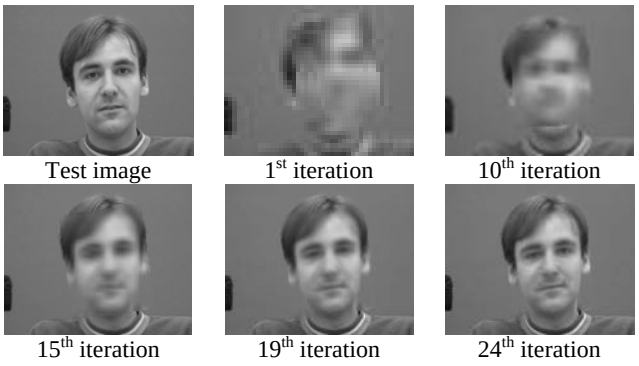


Figure 12: Test image with initial position of the mean model overlapped, and after the 1st, 10th, 15th, 19th and 24th iteration of the search with the active appearance model built for the regular face mesh model.

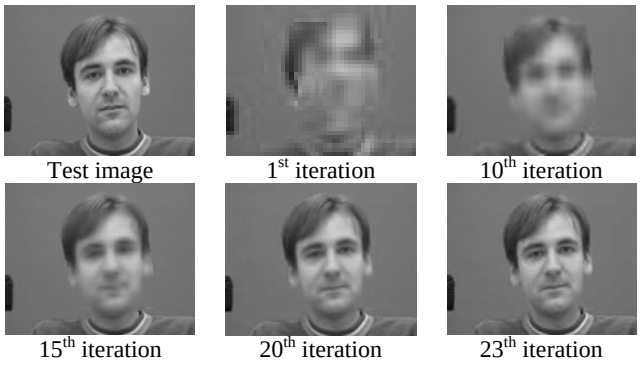


Figure 13: Test image with initial position of the mean model overlapped, and after the 1st, 10th, 15th, 20th and 23th iteration of the search with the active appearance model built for the adaptative face mesh model.

ACKNOWLEDGMENTS

The work presented was partially done in the scope of the project "Segmentation, Tracking and Motion Analysis of Deformable (2D/3D) Objects using Physical Principles", with reference POSC/EEA-SRI/55386/2004, financially supported by FCT - Fundação para a Ciência e a Tecnologia from Portugal.

REFERENCES

Angelopoulou, A. N. and A. Psarrou (2004). *Evaluating Statistical Shape Models for Automatic Landmark Generation on a Class of Human Hands*. Int. Arc. of the Photogrammetry, Remote Sensing and Spatial Information Sciences, Istanbul.

Baker, S. and I. Matthews (2002). *Automatic Construction of Active Appearance Models as an Image Coding Problem*. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1380-1384.

Baker, S. and I. Matthews (2004). *Automatic Construction os Active Appearance Models as an Image Coding Problem*. IEEE Transactions on Pattern Analysis and Machine Intelligence 26: 1380-1384.

Campadelli, P., F. Cusmai and R. Lanzarotti (2003). *A color based method for face detection*. International Symposium on Telecomunications, Isfahan, Iran.

Carvalho, F. J. S. and J. M. R. S. Tavares (2005). *Metodologias para identificação de faces em imagens: Introdução e exemplos de resultados*. Congresso de Métodos Numéricos en Ingeniería 2005, Granada, Espanha.

Cootes, T. F., *Build_aam*, http://www.wiau.man.ac.uk/~bim/software/am_tools_doc/download_win.html, 2004.

Cootes, T. F., *Talking Face*, http://www.isbe.man.ac.uk/~bim/data/talking_face/talking_face.html, 2004.

Cootes, T. F., G. J. Edwards and C. J. Taylor (1998). *Active Appearance Models*. Proc. European Conference on Computer Vision.

Cootes, T. F. and C. J. Taylor (1992a). *Active Shape Models - 'Smart Snakes'*. Proc. British Machine Vision Conference, Leeds.

Cootes, T. F. and C. J. Taylor (1993). *Active Shape Model Search using Local Grey-Level Models: A Quantitative Evaluation*. British Machine Vision Conference, BMVA Press**:** 639/648.

Cootes, T. F., C. J. Taylor, D. H. Cooper and J. Graham (1992). *Training Models of Shape from Sets of Examples*. Proc. British Machine Vision Conference, Leeds.

Cootes, T. F., C. J. Taylor and A. Lanitis (1994). *Active Shape Models: Evaluation of a Multi-Resolution Method for Improving Image Search*. British Machine Vision Conference, BMVA.

Hamarneh, G., *ASM (MATLAB)*, http://www.cs.sfu.ca/~hamarneh/software/code/asm.zip, 1999.

Hicks, Y., D. Marshall, R. R. Martin, P. L. Rosin, M. M. Bayer and D. G. Mann (2002). *Automatic Landmarking for Building Biological Shape Models*. Int. Conf. Image Processing, Rochester, USA 2: 801-804.

Hill, A. and C. J. Taylor (1994). *Automatic Landmark Generation for Point Distribution Models*. Fifth British Machine Vision Conference, England, York, BMVA Press.

Lim, J. S. (1990). *Two-Dimensional Signal and Image Processing*, PTR Prentice Hall.

Stegmann, M. B. and D. D. Gomez, *Hand images*, http://www2.imm.dtu.dk/pubdb/views/publication_details.php?id=403, 2002.